

---

Département des systèmes  
agroalimentaires et ruraux  
CIRAD-SAR

**RAPPORT DE MISSION  
SENEGAL**


04/07/1994 au 23/07/1994

**FORMATION A LA STATISTIQUE  
DESCRIPTIVE SOUS LISA**

**(Statistique uni et  
multidimensionnelle)**

Michel PASSOUANT

CIRAD-SAR/N°77.94  
Septembre 1994



## **SOMMAIRE**

I. Le contexte . . . . .	1
II. Banque de données SAED . . . . .	2
A. Evaluation de la banque de données . . . . .	2
B. Système d'information géographique . . . . .	2
C. Constitution des fichiers de données . . . . .	3
III. Conditions de la formation . . . . .	3
IV. Contenu de la formation . . . . .	4
A. Informatique de base . . . . .	4
B. Définitions et concepts . . . . .	4
C. Conception des enquêtes et des questionnaires . . . . .	5
D. Gestion des données avec le logiciel LISA . . . . .	6
E. Statistique descriptive univariée . . . . .	7
F. Sondages : théorie et pratique . . . . .	7
G. Statistique multivariée . . . . .	8
V. Evaluation . . . . .	9
VI. Conclusion . . . . .	10

### **ANNEXES :**

Annexe 1	Termes de références du stage de formation
Annexe 2	Annonce du stage à la SAED
Annexe 3	Planning de la formation
Annexe 4	Liste des participants
Annexe 5	Liste des supports de cours
Annexe 6	Introduction à Windows : transparents présentés
Annexe 7	Structure du fichier de données OPAMIST.GST
Annexe 8	Exemple de structuration d'un questionnaire en niveaux
Annexe 9	Histogramme et diagramme en bâtons
Annexe 10	Nuage de point multidimensionnel

Dans le cadre du marché à clientèle n° 92/001400, le ministère français de la coopération a demandé au CIRAD-SAR de réaliser une mission d'appui à la SAED pour former les cadres de cette société à la conception des enquêtes en milieu rural et au traitement des données collectées.

Cette mission, effectuée par M. Passouant, responsable du laboratoire INFOBIOM du CIRAD-SAR, a été réalisée du 04/07/94 au 23/07/94, au siège de la SAED, à Saint Louis

## **I. Le contexte**

La SAED doit, dans le cadre des missions qui lui ont été assignées, remplir le rôle "d'observatoire socio-économique du développement rural dans la vallée du Fleuve Sénégal".

Pour mener à bien cette mission, la SAED a mis en place depuis 1990, avec l'appui technique du CIRAD-SAR et financier du Ministère Français de la Coopération une banque de données de suivi des aménagements et des acteurs du développement dans la vallée du Fleuve Sénégal.

De plus une série d'enquêtes et d'études ont été engagées, toujours avec l'appui de la Coopération française, pour produire des statistiques et des références technico-économiques.

La banque de données, en particulier, a conduit la SAED à accumuler une grande masse d'informations qu'il faut d'une part traiter et mettre en forme, et d'autre part qui constitue une base de départ pour de multiples études nécessitant un recueil d'informations complémentaires, par enquêtes notamment.

La bonne réalisation de ce programme d'études nécessite d'une part une bonne connaissance des méthodes de recueil, de mise en forme et de traitement statistique des données, la disposition et la maîtrise de l'outil informatique pour réaliser calculs, graphiques et documents, mais aussi d'autre part l'acquisition par tous les cadres de la DPDR (Direction de la Planification et du Développement Rural) d'une approche rationnelle, homogène et relevant d'une conception commune du développement rural dans la vallée.

A cet effet, un programme de formation a été conçu spécialement pour la SAED, inspiré du cycle ISSEA<sup>1</sup> que le CIRAD-SAR organise régulièrement tous les ans à Montpellier, au printemps, depuis 1991. L'enseignement a été adapté au contexte de la SAED et s'appuie sur des données propres à cet organisme, en particulier extraites de la banque de données, et utilise les logiciels choisis par la SAED :

---

<sup>1</sup> Informatique et Statistique pour le Suivi d'Exploitations Agricoles



Winword et Excel en bureautique, LISA (du CIRAD-SAR) en gestion et traitement de données statistique.

La mission, qui a duré trois semaines, a comporté 2 grandes parties :

- \* 1 semaine consacrée à un bilan de l'information disponible dans la banque de données et à la constitution, par extraction de la banque, de jeux de données pour la formation. A cette occasion, il a été effectué une analyse des problèmes rencontrés sur le plan informatique (retouches de programmation, corrections diverses..) et à la mise en relations de la banque de données et de l'atlas cartographique des aménagements.

- \* 2 semaines consacrées à la formation proprement dite selon le programme de travail ci-après décrit.

## **II. Banque de données SAED**

### **A. Evaluation de la banque de données**

La banque de données a fait l'objet de 6 missions du CIRAD-SAR, à ce jour entre mai 1990 et octobre 1993.

La première semaine a été l'occasion de faire un point du fonctionnement de la banque de données, d'analyser les améliorations et modifications apportées par la SAED et de corriger les quelques dysfonctionnements identifiés.

Tout ce travail a été réalisé avec l'administrateur de la banque de données, M. Abdourahmane KANE, de la DPDR.

### **B. Système d'Information Géographique**

La mise en place d'un Système d'Information Géographique (S.I.G.) est prévue dans le marché de mise à disposition d'experts auprès de la SAED et de travaux de télédétection.

Il s'agit de fournir à la SAED un outil lui permettant de maîtriser la gestion de l'espace agricole à partir d'une cohérence globale intégrant contraintes hydrauliques et politique d'attribution de terres par les communautés rurales.

Pour préparer la fusion de l'atlas cartographique des aménagements, réalisé à partir de la photo-interprétation des images SPOT, avec la banque de données de suivi des aménagements, la SAED accueille un stagiaire du CNEARC.

En complément de cette évaluation de la banque, un appui plus spécifique a été apporté à M. François Comby, stagiaire du CNEARC-ESAT 1 auprès de la SAED,



qui travaille sur le SIG du secteur de Grande Digue Tellel Kassack (GDTK) utilisant le fond de carte communiqué par la société GEOSYS, sous-traitante de BDPA, dans le cadre de l'opération Atlas des aménagements.

Le premier travail de François Comby a constitué en une correction fine du fond de plan (en version de travail de fin mai) en réalisant en particulier :

- effacement de certains traits parasites du dessin
- la création d'une couverture aménagement
- avec affectation des codes UMV aux polygones existants et complément des polygones absents
- le contrôle et la correction des codes aménagements et UMV
- la séparation de la couche routes, pistes et réseau hydraulique
- la jonction avec la banque de données, au travers des fichiers aménagement et UMV.

Ce travail de reprise du fond de plan va permettre la production de cartes thématiques ou de synthèses à partir de requêtes de la banque de données spatialisées soit au niveau aménagement, soit U.M.V..

Le mémoire de François Comby, soutenu au CNEARC le 13 septembre 1994, retrace les principaux résultats obtenus.

### **C. Constitution des fichiers de données**

La réalisation de ce stage spécifique s'accompagne nécessairement de l'adaptation des jeux de données appartenant à la SAED. La banque de données et sa masse d'information accumulée fournit un support idéal à l'apprentissage du logiciel LISA (qui est capable de relire les fichiers de la banque de données) pour sa partie gestion de données, statistique élémentaire et sondage.

A cet effet, nous avons choisi de travailler sur la délégation de Dagana. Les 695 aménagements de cette délégation ont été retenus, et ont été rapprochés des données de surface de l'ensemble de leurs U.M.V. ainsi que des caractéristiques de leur O.P. gestionnaire.

## **III. Conditions de la formation**

La formation s'est déroulée dans la salle de conférence de la direction générale de la SAED à Saint-Louis. Pour la circonstance, une dizaine de micro-ordinateurs avaient été rassemblés, et il a été possible de disposer d'un rétro-projecteur.

Pour chacun des grands thèmes enseignés, un support de cours a été remis, et il sera trouvé en annexes la liste des documents fournis par le CIRAD-SAR.

Toutes les méthodes enseignées ont fait l'objet d'une mise en oeuvre concrète par les stagiaires sur les ordinateurs installés dans la salle. Il y avait en moyenne un ordinateur pour deux personnes.

La dernière version du logiciel LISA a été utilisée et ce fut l'occasion pour la SAED de mettre à jour ses logiciels de statistique.

La présence à Saint-Louis des ordinateurs de délégations de Dagana, Podor et Matam fut aussi l'occasion pour l'administrateur de la banque de données, Abdourahmane KANE, de mettre à niveau les versions de la banque de données des 3 délégations.

#### **IV. Contenu de la formation**

La formation couvre toutes les étapes de la mise en place de l'enquête à l'exploitation des données et l'interprétation des résultats.

##### **A. Informatique de base - 1 jour**

A cet effet, le stage commence par une remise à niveau informatique : l'ordinateur est en effet aujourd'hui un incontournable outil de stockage, d'analyse et de valorisation de la masse de données recueillies au cours d'une enquête.

Dans ce premier jour, les principes de base de micro-informatique sont rappelés, depuis l'architecture matérielle jusqu'aux systèmes MS/DOS et WINDOWS, en insistant sur les précautions et les principes qui doivent guider l'utilisation de fichiers (organisation de disques, copies de sécurités...).

Après ces rappels, une mise à niveau bureautique a été faite pour permettre une rédaction et mise en forme graphique des résultats d'enquête.

A cette partie de la formation ont aussi assisté les secrétaires de la DPDR et ce fut pour elles l'occasion d'acquérir quelques concepts de micro-informatique. Cela leur aidera dans leur activité de secrétariat en permettant une meilleure utilisation des logiciels de bureautique et de traitement de texte en particulier.

##### **B. Définitions et concepts - 0,5 jour**

La mise en place de la banque de données a été l'occasion d'harmoniser au sein de la SAED les divers concepts utilisés (aménagement, surface nette irrigable initiale, exploitable, abandonnée, en extension, acteurs du développement..) ainsi que les nomenclatures associées à diverses variables qualitatives (type d'aménagement, de pompes, raisons d'abandon de terre, type d'O.P....).

Ce travail, fruit d'un consensus à établir entre tous les agents de la SAED, et en particulier de la DPDR, a été repris pendant une demi-journée à l'occasion de la



réunion pour cette formation des différents chargés d'enquêtes et agents de suivi-évaluation des délégations.

L'objectif poursuivi ici est d'harmoniser au niveau de la SAED les concepts et le vocabulaire utilisé pour :

- assurer la cohérence des informations collectées de manière indépendantes
- permettre des comparaisons entre enquêtes SAED
- établir des définitions claires communes à toute la SAED
- diffuser à l'extérieur les concepts ici définis

et par là conforter l'unité et la fiabilité des études SAED.

### **C. Conception des enquêtes et des questionnaires - 1 jour**

Les enquêtes, vastes opérations de recueils lourds sont généralement décevantes quant à la qualité des données rassemblées, des analyses qui peuvent être menées et des conclusions pour apporter réponse aux questions qui ont suscité la mise en place de l'enquête.

La collecte des données est en réalité un processus d'échanges de messages et d'informations entre d'une part le chargé d'enquête et ses enquêteurs, puis les enquêteurs et les individus enquêtés et enfin restitution des informations à l'analyste, par l'intermédiaire d'une saisie informatique et de calculs statistiques.

Chacun de ces maillons, points de discontinuité dans le circuit d'information, peut être à l'origine d'erreurs dues :

- une mauvaise compréhension des objectifs de l'enquête
- une interprétation différente du message entre l'émetteur et le récepteur
- des fautes d'inattentions ou de retranscription

et donc un soin tout particulier doit être apporté pour définir clairement les concepts utilisés et la formulation des questions afférentes d'une part, et contrôler tous le processus de collecte et codage des données.

Cette étape, bien souvent négligée et abordée trop rapidement est en réalité fondamentale pour la qualité des données collectées :

- l'objet de l'enquête doit être clairement posé, et à partir de cela la définition de la population sujet est définie très précisément.
- la population ainsi définie peut cacher des niveaux d'observations différents qu'il faut savoir identifier (exploitation/parcelle/campagne ou ménage/membres du ménage ou troupeau/tête de bétail...).
- les informations, ou variables observées, ou questions à poser sont ensuite définies, en prenant séparément les variables quantitatives, pour lesquelles seront définies unités de mesure, des variables qualitatives dont on construit les nomenclatures.



- il faut ensuite structurer ceci en un questionnaire qui sert à la collecte et la saisie informatique. Ce questionnaire doit faire impérativement l'objet d'explications approfondies auprès des enquêteurs et doit être testé avant d'être généralisé à l'ensemble de l'échantillon.
- l'échantillon, dans le cas d'enquêtes par sondage, doit être sélectionné avec soin, et selon une méthodologie en rapport avec les données disponibles, les objectifs de nature de l'information à rassembler, sa précision et sa qualité.
- le processus de collecte doit faire l'objet de contrôles permanents et multiples pour identifier au plutôt les données douteuses et les corriger par un retour sur le terrain.
- la saisie informatique est enfin l'objet d'une relecture, ou de contrôles par coups de sondes ou calculs statistiques avant d'aborder les traitements et l'analyse.

#### **D. Gestion des données avec le logiciel LISA - 2 jours**

Le logiciel LISA, développé par le CIRAD-SAR, est utilisé par la SAED depuis 1990 avec satisfaction. Il fournit un support complet pour assurer la gestion et le traitement statistique des données d'enquêtes, en particulier dans le domaine du développement rural.

Le logiciel a été entièrement parcouru en abordant successivement :

- présentation générale et installation
- notion de fichier : structure, champs, valeurs manquantes, à vérifier
- saisie, correction, impression de fichiers
- calculs de champs d'un fichier
- recodage de champs d'un fichier
- effacement d'enregistrements en distinguant logique et physique
- saisie par formulaire
- impression brute ou avec formulaire
- opérations sur fichiers : fusions, extraction, agrégation (pour changement de niveaux).
- importation/exportation, en particulier récupération de fichiers de la banque de données

Durant ces journées, le fichier support pour la formation identifié durant la première semaine de mission a été reconstitué par les stagiaires en faisant successivement :

- une importation de fichier Foxpro de description des aménagements, des U.M.V. et de O.P.
- des modifications de structure (en particulier définition de nomenclatures de variables qualitatives) de ces fichiers.
- l'introduction de champs calculés.
- l'agrégation de fichiers (pour passer des U.M.V. aux aménagements)
- la fusion horizontale de fichiers...

Le fichier ainsi constitué a été nommé OPAMIST.GST, dans la logique de LISA.

### **E. Statistique descriptive univariée - 2 jours**

Si les notions de statistiques élémentaires, telles que moyenne, écart-type.. sont bien connues il est néanmoins nécessaire de les reprendre pour en faire sentir les effets pervers cachés et aider à l'utilisation des résultats fournis par le logiciel pour analyser et éviter des erreurs d'interprétation en isolant soit des individus extrêmes, soit en identifiant des sous-populations bien marquées :

- sensibilité de la moyenne aux valeurs extrêmes
- utilisation conjointe de la moyenne, de la médiane et du mode
- notion de dispersion à partir de l'écart-type, du coefficient de variation et des quartiles
- représentations graphiques : diagrammes en bâtons, histogrammes, nuages de points
- notion de liaison et mesure du degré de la relation selon les natures de variables : tableaux croisés et analyse des contributions des cellules, des lignes ou colonnes, rapport de corrélation, coefficient de corrélation..

Pour cette partie ont été abordées successivement :

- Définitions et vocabulaire
- Statistiques de comptage
- Statistiques de tendance centrale et de dispersion
- Représentations graphiques

et leur mise en oeuvre sous LISA : activation des programmes ; lecture et interprétation des résultats.

Les traitements ont été réalisés sur le fichier OPAMIST.GST construit ci-dessus et ont permis d'analyser les caractéristiques des aménagements de la délégation de DAGANA. A cette occasion certains aménagements aux caractéristiques extrêmes ont été identifiés, isolés et retirés du fichier pour créer des fichiers OPAMIST1, 2 et 3 sur lesquels les paramètres statistiques ont été recalculés.

Un accent tout particulier a été porté sur les imperfections des fonctions statistiques des tableurs et à l'attention à porter sur l'interprétation des résultats numériques et graphiques.

### **F. Sondages : théorie et pratique - 1 jour**

A l'issue des résultats d'une enquête, il se pose généralement la question d'une part de la représentativité de l'échantillon étudié et d'autre part de la généralisation des conclusions à l'ensemble de la population de l'étude.

La théorie des sondages permet de construire des échantillons représentatifs de la population et fournit des moyens d'extrapoler les résultats obtenus à l'ensemble de la population mère. Dans ce cadre, il a été abordé :



- les notions de recensement et de sondage.
- les divers types de sondages : échantillonnage et méthodes d'estimation associées.
- échantillonnage aléatoire (simple, avec ou sans remise, stratifié représentatif et optimal, sondage en grappe).
- échantillonnage non aléatoire, méthode des quotas.
- pratique de l'échantillonnage et de l'estimation.

Pour illustrer cette partie du cours il a été utilisé un logiciel spécifique de Sondage élaboré par G. Francillon, du CNEARC, qui permet de mettre en oeuvre différentes méthodes de tirages d'échantillons à partir d'un fichier à la norme LISA. En particulier, il est ainsi possible de bien mettre en évidence les fluctuations d'échantillonnage et de sensibiliser au caractère aléatoire de l'échantillon étudié dans le cadre d'une enquête.

Les stagiaires ont été formés à l'utilisation de ce logiciel et les exercices ont été faits en utilisant le fichier OPAMIST.GST décrit ci-dessus.

La banque de données mise en place par la SAED fournit plusieurs base de sondage pour construire des échantillons aléatoires : aménagements, UMV ou acteurs du développement. Et pour chacune d'entre elles, diverses variables de stratification peuvent être utilisées en fonction des objectifs de l'enquête.

### **G. Statistique multivariée - 2 jours**

La gestion de la banque de données ou les opérations d'enquêtes conduisent au recueil d'un grand nombre d'informations différentes sur les mêmes individus. La statistique traditionnelle univariée ou bivariée n'exploite que très mal ces données.

L'analyse des données offre, par la prise en compte simultanément d'un grand nombre de variables, la possibilité de mettre en évidence d'une part des relations entre les variables, et d'autre part de constituer des groupes d'individus partageant des caractéristiques communes sur l'ensemble des critères d'observation.

Les méthodes abordées sont classées en deux familles : analyses factorielles qui produisent des cartes ou graphiques factoriels et les méthodes de classification. Il a été vu successivement :

- L'A.C.P., analyse en composantes principales, réservée à des variables quantitatives et au travers de laquelle les notions de nuage de point multidimensionnels, d'inertie attachée, d'axe factoriel et de qualité de représentation ont été formalisées.
- L'A.F.C., analyse factorielle des correspondances, sur tableau de contingence a été replacée dans le contexte formel mis en place pour l'A.C.P. Ensuite a été introduite, par généralisation, l'A.F.C.M. ou analyse factorielle des correspondances multiples, sur tableau disjonctif complet construit à partir d'un tableau de données individus\*variables dans lequel



toutes les variables sont qualitatives.

- Les Classifications ont été vues rapidement, principalement hiérarchiques fondées sur le moment d'ordre 2. Leur utilisation en complément des méthodes factorielles, ainsi que les outils d'aide à l'interprétation des classes ont été exposés.

Le temps imparti à la formation était limité. Compte tenu du caractère nouveau de ces méthodes pour la majorité des auditeurs, et pour ne pas rajouter une complexité due aux jeux de données soumis aux analyses, il a été utilisé des données issues des exemples classiques rassemblés dans la littérature statistique. Ces jeux de données connus pour leurs simplicités et leurs bonnes adéquations aux méthodes présentées, sont de petites tailles et concernent des thèmes extrêmement généraux connus de tous (races de chiens, structure de consommation alimentaire...).

Il reste aux stagiaires à conforter les connaissances acquises à des données plus personnelles issues de leurs travaux au sein de la SAED.

## **V. Evaluation**

A l'issue de la formation, une séance d'évaluation a été organisée rassemblant 15 des stagiaires.

Après un tour de table complet, il est ressorti que, malgré la très forte hétérogénéité de l'assistance, tout le monde considère cette formation comme très profitable et pense trouver très prochainement une application dans son activité professionnelle.

Il est bien évident, que les jugements individuels sont tempérés par les compétences initiales, et que le niveau ciblé était forcément un compromis. Si certains ont pu être un peu dépassés par la dernière partie au formalisme mathématique plus abstrait, par contre d'autres ont regretté qu'il n'est pas été possible d'approfondir plus longuement les méthodes multidimensionnelles. Une présentation avec un développement mathématique plus poussé a même été demandé.

En tout état de cause notre objectif était de donner une base de départ pour aborder les analyses d'enquêtes. La mise en pratique est une nécessité évidente, et la maîtrise des outils statistiques et informatique ne peut être que le résultat d'une expérience acquise par le traitement de données personnelles.

## VI. Conclusion

L'outil de statistique, essentiellement descriptif présenté au cours des journées a pour soucis principal de décrire la diversité.

La théorie des sondages veut construire un échantillon couvrant toute la diversité de la population mère pour la décrire de façon la plus fidèle possible.

La statistique descriptive veut d'une part résumer cette diversité, mais aussi la restituer et en décrire les grands traits structurant au travers de la notion d'inertie et de reconstruction de l'inertie.

Toutes ces méthodes ont fait l'objet d'implémentations informatiques dans des logiciels largement répandus, qu'il s'agisse de grands produits commerciaux (SAS, SPSS, SysStat...) que de produits moins commerciaux et plus ciblés sur le domaine rural (Statitcf, Cstat, LISA...).

Cependant on se rend compte que les utilisateurs restent très circonspects devant la richesse et la complexité des traitements statistiques qui leur sont proposés.

En réalité ces logiciels sont bien souvent sous-utilisés, voire abandonnés au profit de tableurs, du fait de l'insuffisance de connaissances statistiques de leurs clients.

Il est donc fondamental d'assurer des formations statistiques en complément, voire comme unique vecteur, de la diffusion d'un logiciel de statistique. Notre expérience de 4 années dans cette activité, avec LISA comme support informatique, ne peut que nous conforter dans cette idée.

Ces actions de formations statistiques vont se poursuivre et joueront un rôle fondamental dans la diffusion du nouveau logiciel, WINSTAT, que le CIRAD s'apprête à lancer dans ce domaine, malgré l'apport de l'environnement WINDOWS et tous les efforts d'amélioration de l'ergonomie réalisés.

# ANNEXES

Annexe - 1 : Termes de références du stage de formation

Annexe - 2 : Annonce du stage à la SAED

Annexe - 3 : Planning de la formation

Annexe - 4 : Liste des participants

Annexe - 5 : Liste des supports de cours

Annexe - 6: Introduction à WINDOWS : transparents présentés

Annexe - 7 : Structure du fichier de données OPAMIST.GST

Annexe - 8 : Exemple de structuration d'un questionnaire en niveaux

Annexe - 9 : Histogramme et diagramme en bâtons

Annexe - 10 : Nuage de point multidimensionnel



## Annexe - 1 : Termes de références du stage de formation

### **S.A.E.D.**

**Direction de la Planification et du  
Développement Rural**

#### **Termes de référence pour la réalisation d'un stage de formation des cadres de la SAED à la conception des enquêtes en milieu rural et au traitement des données collectées**

##### **1 - Cadre général et objectif de la mission.**

Parmi les missions assignées à la SAED figure celle "d'observatoire socio-économique du développement rural dans la Vallée du Fleuve Sénégal". Pour assumer cette fonction, la SAED doit mener un certain nombre d'opérations de suivi et d'évaluation basées sur des traitements de données collectées pour la plupart par enquêtes auprès des différents intervenants des filières concernées, et en particulier auprès des exploitations agricoles et des organisations paysannes.

La SAED s'est engagée, avec l'appui du Ministère français de la coopération (convention FAC/72/CD/92), dans une série d'enquêtes et études qui doit lui permettre de produire les statistiques et les références technico-économiques nécessaires à une bonne planification du développement agricole de la Vallée à travers la mise en oeuvre du PDRG.

Pour mener à bien ces opérations, la SAED dispose de cadres de formations diverses. Or, il est nécessaire que les concepts, les méthodes, les outils (logiciel informatique) et une partie au moins des traitements soient communs et homogènes, entre les différentes études et entre les délégations. Par ailleurs, des rappels et remises à niveau sont nécessaires pour s'assurer de l'opérationnalité de tous.

Il s'agit donc d'organiser un stage de 2 semaines pour une dizaine de cadres à la SAED sur la conception, la réalisation et le traitement d'enquêtes en milieu rural.

Pour préparer et réaliser ce stage, la SAED souhaite bénéficier de l'appui d'un expert.

En raison de l'expérience du CIRAD-SAR en général, et de celle M. PASSOUANT en particulier dans ce domaine (avec notamment la réalisation annuelle d'un "stage informatique et statistique pour le suivi des exploitations agricoles" - ISSEA à Montpellier), en raison de l'utilisation du logiciel LISA (produit du CIRAD-SAR) par quelques cadres de la SAED et des satisfactions qu'il apporte, la SAED souhaite bénéficier de l'appui de M. PASSOUANT pour la préparation et la réalisation de cette formation.

##### **2 - Description du stage.**

L'objectif du stage est de compléter la formation des cadres de la SAED (DPDR, Délégations et autres services) en charge des opérations de suivi, d'évaluation et d'études dans les domaines de la conception et la mise en oeuvre d'enquêtes, de la gestion des données avec un outil informatique et de l'exploitation de ces données.

Pour l'exploitation des données, les rappels et compléments de formation concerneront en priorité les statistiques descriptives (moyennes, écart-type, histogrammes, tableaux croisés, régression, etc...) mais devront offrir une ouverture vers les statistiques multivariées (ACP et AFC) pour les personnes plus avancées.

Par ailleurs, si ce stage s'avère nécessaire pour mettre à niveau des cadres de formation diversifiés, il doit également permettre l'adoption de méthodes et d'outils communs pour la mise en oeuvre d'une démarche homogène au sein de la SAED.

La SAED a choisi d'utiliser le logiciel LISA pour gérer et traiter les données d'enquêtes et les logiciels EXCEL et WINWORD pour présenter les résultats (tableur-grapheur et traitement de texte). Ce stage permettra de généraliser l'utilisation de ces outils (rappels et compléments).

Le stage sera avant tout "pratique". Les théories seront présentées brièvement pour laisser le maximum de temps aux travaux pratiques qui seront réalisés avec des données collectées par la SAED (et notamment les données de la banque sur le suivi des aménagements hydro-agricoles et des organisations paysannes et les données des enquêtes d'exploitations agricoles en cours de traitement).

Les thèmes abordés durant le stage sont présentés ci-dessous, avec une proposition de durée pour chacun d'eux :

Rappels informatiques concernant MS-DOS, WINWORD et EXCEL	1 jour
Définition des différents concepts utilisés (exploitations agricoles, organisations paysannes, aménagements hydro-agricoles, UMV, etc.)	1 jour
Conception des enquêtes et réalisations des questionnaires	1 jour
Gestion des données sous LISA	2 jours
Théorie et pratique des sondages	1 jour
Statistiques descriptives (4 jours pour les moins avancés)	2 jours
Introduction à l'analyse de données pour les plus avancés	2 jours

Au total, le stage durera 2 semaines. La SAED mettra à la disposition de M. Passouant les données de la SARO qu'il utilisera comme support de formation.

La Répartition des rôles pour l'organisation et l'encadrement du stage entre la SAED et l'expert est la suivante :

Le stage est organisé par la SAED (locaux adaptés, micro-ordinateurs et périphériques, tableaux).

La SAED dispose déjà des logiciels LISA, EXCEL et WINWORD.

Les données pour les travaux pratiques seront fournies par la SAED en début de stage.

Le CIRAD-SAR fournira un manuel d'utilisation de LISA par stagiaire (gestionnaire de données et statistiques).

La SARO/DPDR appuiera M. Passouant pour les deux premières journées : rappels informatiques sur EXCEL et WINWORD et définition des concepts.

### 3 - Réalisation de la mission d'appui

Les tâches à réaliser par l'expert sont les suivantes :

- fourniture des 2 manuels de LISA (statistiques et gestion de données) en 10 exemplaires chacun ;
- préparation des supports de formation à partir des données fournies par la SAED (3 jours) ;
- organisation pratique du stage (1 jour) ;
- réalisation de la formation (2 semaines) ;
- rédaction d'un compte-rendu de stage à fournir en 5 exemplaires à la SAED.

La durée de la mission est de 20 jours au Sénégal.

La mission devra être réalisée durant le courant du mois de Juin ou de Juillet 94.

*Stage de formation sur la  
conception, la réalisation et le  
traitement d'enquêtes en  
milieu rural à l'intention des  
cadres de la SAED.*

*Durée : du Lundi 11/07/94 au Vendredi 22/07/94*

*Lieu : Salle de conférence de la Direction Générale*

*Animateur : Monsieur Passouant du CIRAD/SAR*



Annexe - 3 : Planning de la formation

# Stage de formation des cadres de la DPDR à la conception, la réalisation et l'exploitation d'enquêtes en agriculture irriguée

Saint-Louis du 11 au 22 Juillet 1994

Date	Heures	Thème	Intervenant
LUNDI 11/07	Matin	<u>Rappels de micro-informatique</u> : MS/DOS, Windows	M.Passouant
	Après-midi	<u>Rappels de bureautique</u> : Winword et Excel	J.F.Bélières
MARDI 12/07	Matin	<u>Définition des concepts</u> utilisés pour des enquêtes et pour la banque de données : exploitation agricole, OP, aménagement, UMV...	J.F.Bélières
	Après-midi	<u>Conception et réalisation de questionnaire</u> - Définition des objectifs - Définition de la population - Définition des informations à recueillir et des nomenclatures - Structuration du questionnaire, notion de niveau - Informatisation du questionnaire	J.F.Bélières & M.Passouant
MERCREDI 13/07	Matin	<u>Conception et réalisation de questionnaire(suite)</u> - Test du questionnaire - Rédaction du manuel de l'enquêteur et formation des enquêteurs - Organisation de la collecte - Contrôle de la collecte - Lecture, recodage, compléter nomenclatures - Saisie informatique - Contrôle de la saisie questionnaire	M.Passouant & J.F.Bélières
	Après-midi	<u>Gestion des données sous LISA</u> - présentation générale et installation - notion de fichier : structure, champs	M.Passouant
JEUDI 14/07	Matin	<u>Gestion des données sous LISA(suite)</u> - saisie, correction, impression - calculs de champs - recodage	M.Passouant
	Après-midi	<u>Gestion des données sous LISA (suite)</u> - effacement d'enregistrements - saisie par formulaire - impression brute ou avec formulaire	M.Passouant

Date	Heures	Thème	Intervenant
VENDREDI 15/07	Matin	<u>Gestion des données sous LISA (suite)</u> - opérations sur fichiers : fusions, extraction, agrégation. - importation/exportation	M.Passouant
	Après- midi	<u>Statistique descriptive</u> - définitions et vocabulaire - Statistiques de comptage	M.Passouant
LUNDI 18/07	Matin	<u>Statistique descriptive(suite)</u> - Statistiques de tendance centrale et de dispersion - Représentations graphiques	M.Passouant
	Après- midi	<u>Statistique descriptive(suite)</u> : mise en oeuvre informatique et applications	M.Passouant
MARDI 19/07	Matin	<u>Statistique descriptive(suite)</u>	M.Passouant
	Après- midi	<u>Théorie et pratique des sondages</u> - Recensement / Sondage - Divers types de sondages : échantillonnage et estimation	M.Passouant
MERCREDI 20/07	Matin	<u>Théorie et pratique des sondages(suite)</u> - Echantillonnage - Pratique de l'échantillonnage et estimation	M.Passouant
	Après- midi	<u>Introduction à la statistique multidimensionnelle</u> - A.C.P.	M.Passouant
JEUDI 21/07	Matin	<u>Introduction à la statistique multidimensionnelle(suite)</u> - analyse des correspondances	M.Passouant
	Après- midi	<u>Introduction à la statistique multidimensionnelle(suite)</u> - analyse des correspondances multiples - classifications	M.Passouant
VENDREDI 22/07	Matin	<u>Introduction à la statistique multidimensionnelle(suite)</u> - classifications  <u>Evaluation</u>	M.Passouant



**Annexe - 4 : Liste des participants**

**I. DPDR Saint Louis**

- Mesdames
  - Maïmouna DIA
  - Adiaratou DIALLO
  - Anna THIAN
- Messieurs
  - Jean François BELIERES
  - Amadou CISSE
  - Abdou Khadre DIOP
  - Abdourahmane KANE
  - Arona MANE
  - Jean MOREIRA
  - Madia NDIAYE
  - El Hadji Adama TOURE
  - Abdourahmane SY
  - Fatimata SY DIA

**II. Délégation de DAGANA**

- Messieurs
  - Seydou CAMARA
  - Abdou NDAO (stagiaire)

**III. Délégation de PODOR**

- Monsieur
  - Moussa Bocar FOFANA

**IV. Délégation de MATAM**

- Monsieur
  - Oumar Samba THIAM

**V. CNAPTI**

- Monsieur
  - Assane DIOP (Stagiaire)

**VI. Extérieur SAED**

- Madame
  - Oumou KANE
- Messieurs
  - Abdoulaye CAMARA
  - Samba KANTE (ISRA)
  - Laurent LIAGRE (CIRAD)

## Documents remis aux participants

MS/DOS - Laboratoire INFOBIOM - CIRAD-SAR

LISA 3.1 - Gestionnaire - G. Francillon, J.C. Sicard, P. Sada-Tailly - CIRAD/SAR N° 1992/07

LISA 3.1 - Statistique - G. Francillon, J.C. Sicard, P. Sada-Tailly - CIRAD/SAR N° 1992/07

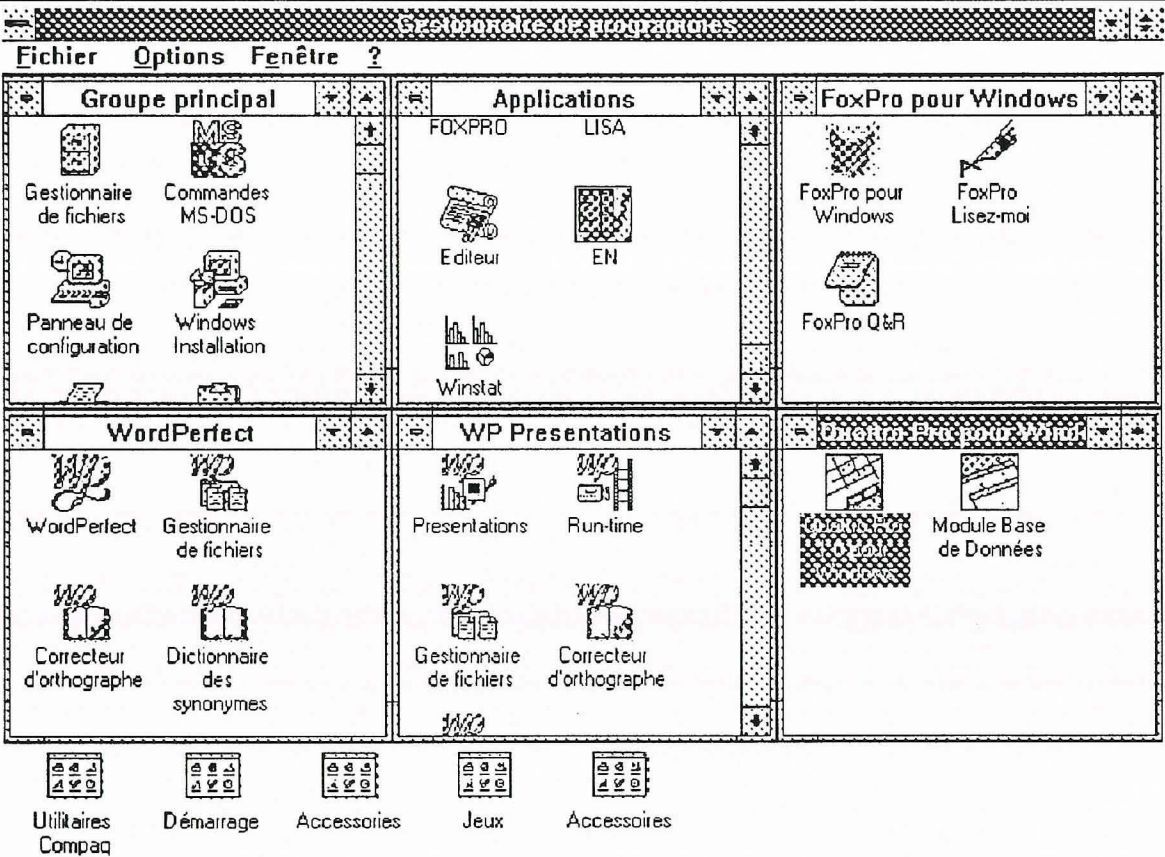
Intérêt de la simulation informatique dans l'étude des techniques de sondage - Tome II Sondage, un logiciel pour l'étude des sondages - G. Francillon - ENSAA & CNEARC Juillet 1992

Méthodes Statistiques sous LISA - I Statistique univariées et Sondages - M. Passouant, C. Troude, R. Lenoir - CIRAD/SAR N° 40/93

Méthodes Statistiques sous LISA - II Statistique multivariée - M. Passouant, C. Troude, R. Lenoir - CIRAD/SAR N° 40/93

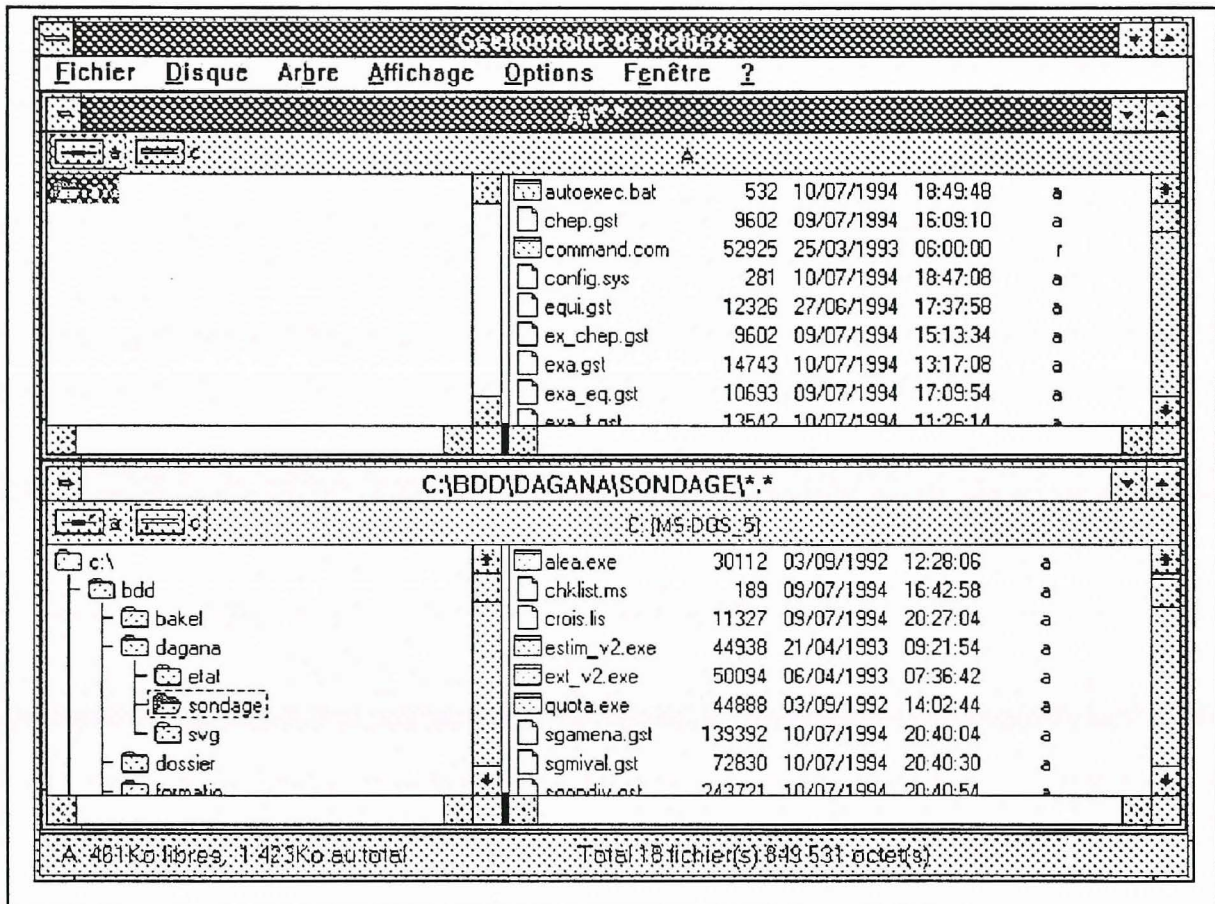


# WINDOWS - Le gestionnaire de programmes



# WINDOWS - Le gestionnaire de fichiers

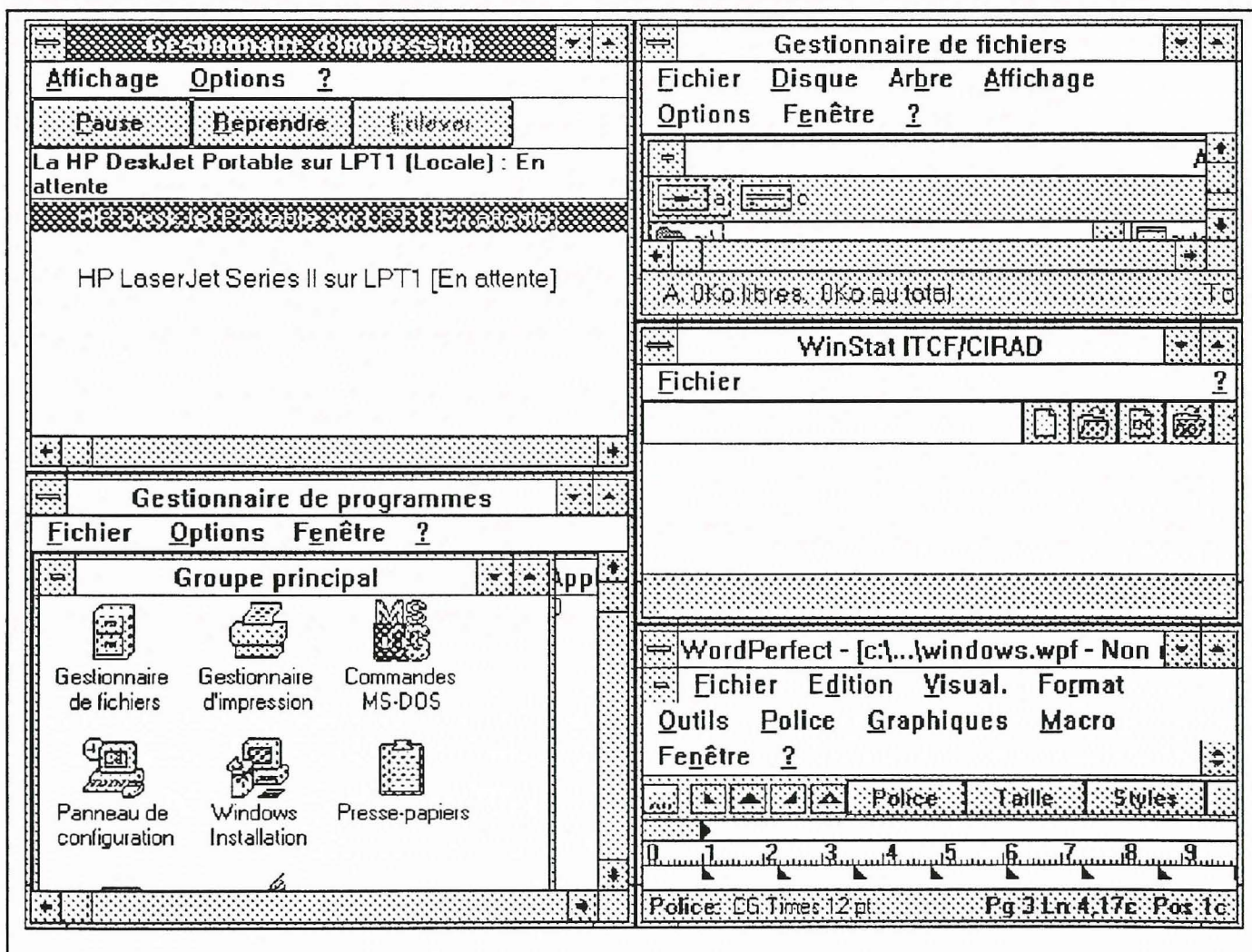
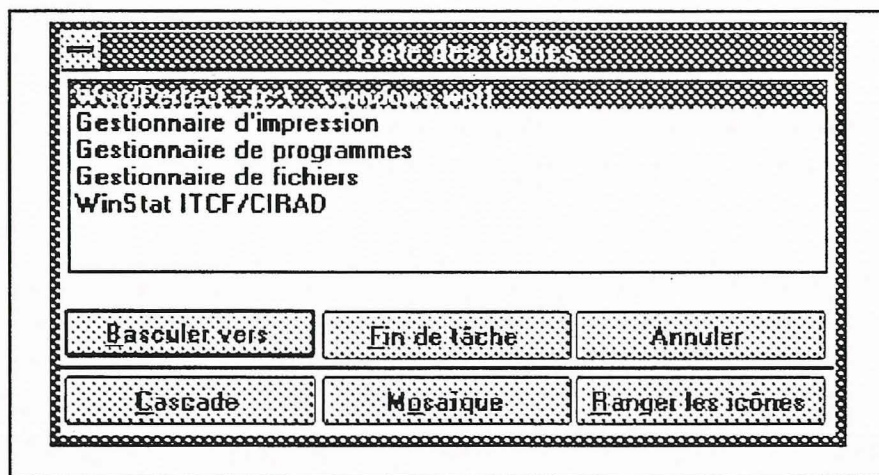
Par le gestionnaire de programme : Icône gestion de fichier du groupe principal





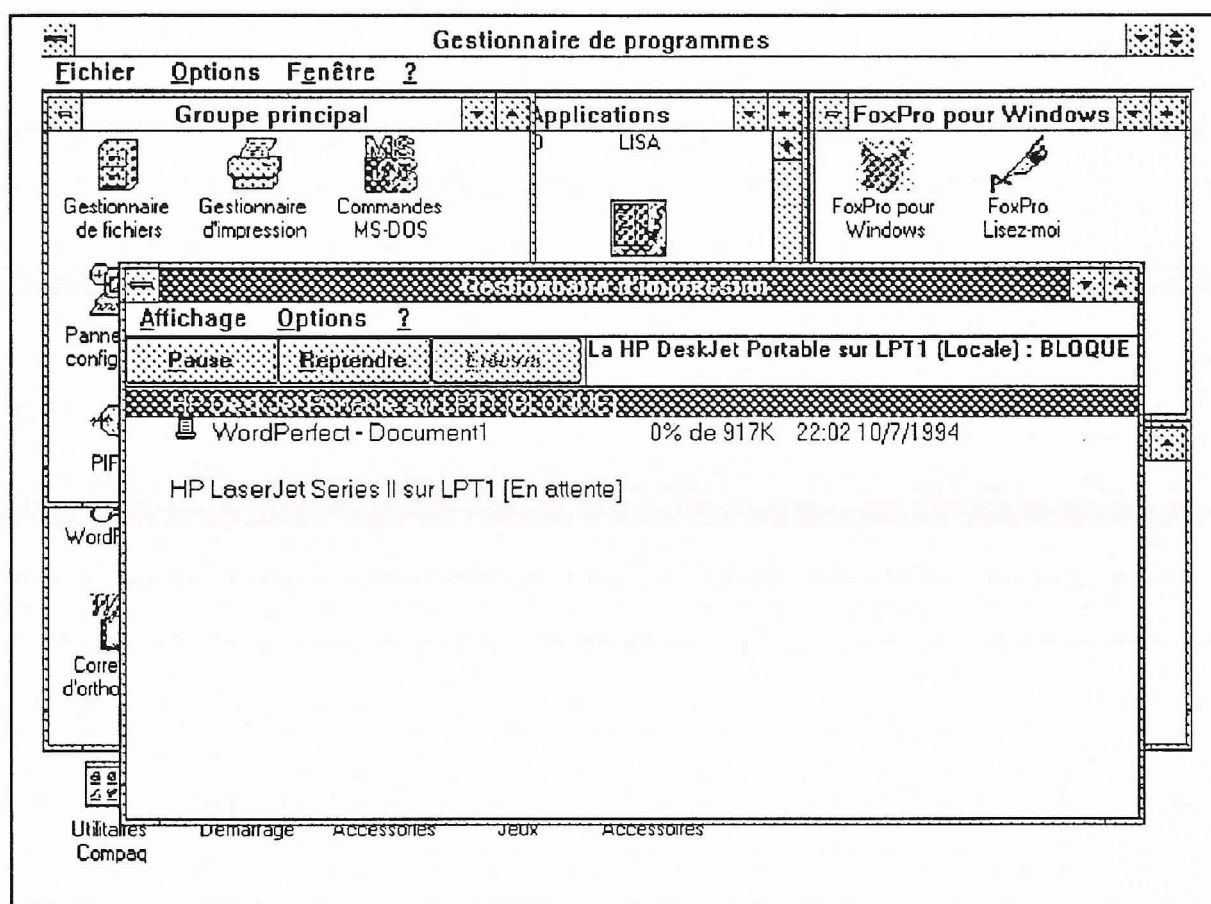
# WINDOWS - Le gestionnaire de tâches

Soit la case système de toute fenêtre WINDOWS, Soit par les touches Alt-TAB



# WINDOWS - Le gestionnaire d'impression

Par le gestionnaire de programme : Icône gestion de d'impression du groupe principal





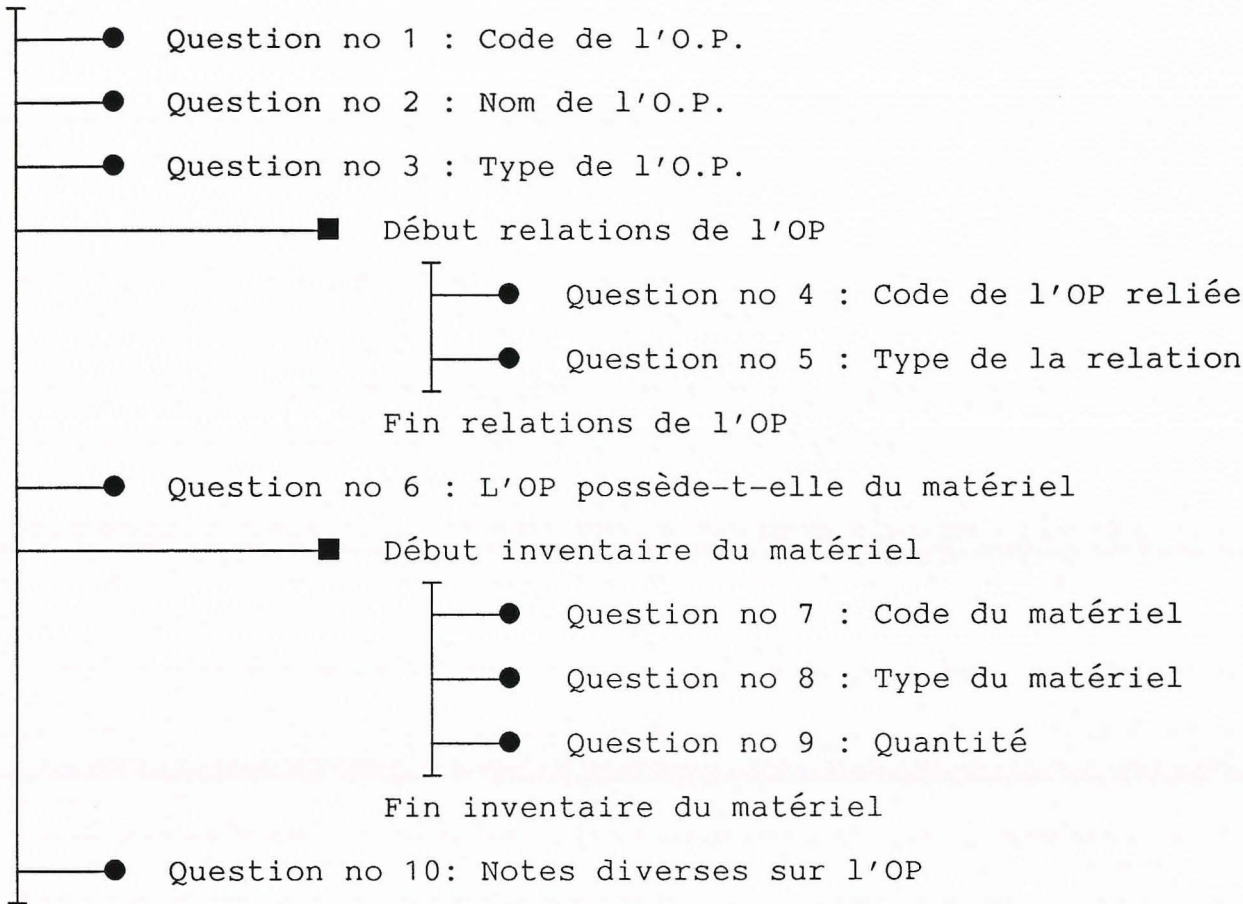
Annexe - 7 : Structure du fichier de données OPAMIST.GST

STRUCTURE DU FICHIER..... C:\BDD\DAKANA\SONDAGE\OPAMIST.GST									
19 champs - 695 enregistrements									
C:\BDD\DAKANA\BDD\DIV.DBF									
Ch. Nom	Nom	a	Type	long	Déc	Modalités	Limites de signification		
1 IDENT	IDENT		C	4					
2 K_S_N_IRR	SIRR		F	8	2	0	[	*	*)
3 K_S_EXTEN	SEXT		F	8	2	0	[	*	*)
4 K_S_ABAND	SADA		F	8	2	0	[	*	*)
5 K_S_REPLO	SEXP		F	8	2	0	[	*	*)
6 CLASS_TAIL	CRXX		Q	2	0	11	[	1,	11)
						Modalité 1 de CLASS_TAIL :	ZERO	ZER	
						Modalité 2 de CLASS_TAIL :	INF5	15	
						Modalité 3 de CLASS_TAIL :	INF10	110	
						Modalité 4 de CLASS_TAIL :	INF20	120	
						Modalité 5 de CLASS_TAIL :	INF30	130	
						Modalité 6 de CLASS_TAIL :	INF40	140	
						Modalité 7 de CLASS_TAIL :	INF50	150	
						Modalité 8 de CLASS_TAIL :	INF80	180	
						Modalité 9 de CLASS_TAIL :	INF120	1120	
						Modalité 10 de CLASS_TAIL :	INF250	1250	
						Modalité 11 de CLASS_TAIL :	SUP250	S250	
7 T_AMENAGE	T_AN		Q	2	0	4	[	1,	4)
						Modalité 1 de T_AMENAGE :	AM_CRAND	AMGR	
						Modalité 2 de T_AMENAGE :	AM_INTER	AMIT	
						Modalité 3 de T_AMENAGE :	AM_PPI	AMPP	
						Modalité 4 de T_AMENAGE :	AM_PP_HORS	AMPH	
8 K_D_AM_VI	K_VI		F	2	0	0	[	*	*)
9 K_D_AM_CA	K_CA		F	2	0	0	[	*	*)
10 K_D_AM_DS	K_DS		F	3	0	0	[	*	*)
11 T_FAC_PIS	T_PA		Q	1	0	4	[	0,	3)
						Modalité 1 de T_FAC_PIS :	ACC_INCONN	ACIC	
						Modalité 2 de T_FAC_PIS :	ACC_BON	ACBO	
						Modalité 3 de T_FAC_PIS :	ACC_MOYEN	ACMO	
						Modalité 4 de T_FAC_PIS :	ACC_DIFFIC	ACDF	
12 T_ENCL_HI	ENCL		Q	1	0	3	[	0,	2)
						Modalité 1 de T_ENCL_HI :	ENCL_INCONN	ENIC	
						Modalité 2 de T_ENCL_HI :	ENCLAVE	ENCL	
						Modalité 3 de T_ENCL_HI :	EN_PAS	ENPA	
13 T_NAT_STA	T_NA		Q	1	0	6	[	0,	5)
						Modalité 1 de T_NAT_STA :	NS_INCONN	NSIN	
						Modalité 2 de T_NAT_STA :	NS_FIXE	NSFM	
						Modalité 3 de T_NAT_STA :	NS_GMP_FLO	NSGF	
						Modalité 4 de T_NAT_STA :	NS_GMP_DR	NSGB	
						Modalité 5 de T_NAT_STA :	NS_GMP_LOC	NSGL	
						Modalité 6 de T_NAT_STA :	NS_GOLAIRE	NSSL	
14 T_ENERGIE	T_EN		Q	1	0	3	[	0,	2)
						Modalité 1 de T_ENERGIE :	EN_INCONN	ENIC	

				Modalité 2 de T_ENERGIE :	EN_ELECTRI ENEL	
				Modalité 3 de T_ENERGIE :	EN_DIESEL ENDE	
15	T_STATION	T_ST	0 2 0 5	(	0,	4]
				Modalité 1 de T_STATION :	TS_INCONNU TSIC	
				Modalité 2 de T_STATION :	TS_INRIC TSIR	
				Modalité 3 de T_STATION :	TS_EXHAUSTE TSSE	
				Modalité 4 de T_STATION :	TS_MIXTE TSIX	
				Modalité 5 de T_STATION :	TS_GMP TSGM	
16	K_NB_MBR	K_NB	F 3 0 0	(	*,	*)
17	TX_ABAND	TXAN	F 8 2 0	(	*,	*)
18	T_OP	TOFA	0 2 0 11	(	0,	10]
				Modalité 1 de T_OP :	TO_INCONNU TOIC	
				Modalité 2 de T_OP :	TO_SECT_VI TOGV	
				Modalité 3 de T_OP :	TO_GROUP TOGR	
				Modalité 4 de T_OP :	TO_GIE TOG	
				Modalité 5 de T_OP :	TO_GIE_EX TOGE	
				Modalité 6 de T_OP :	TO_GIE_PA TOGP	
				Modalité 7 de T_OP :	TO_GIE_DV TOGD	
				Modalité 8 de T_OP :	TO_EXP_PHV TOPR	
				Modalité 9 de T_OP :	TO_FOTER TOFO	
				Modalité 10 de T_OP :	TO_FEDE_OP TOFD	
				Modalité 11 de T_OP :	TO_AUTRE TOAU	
19	T_OP_PREC	T_OP	0 2 0 11	(	0,	10]
				Modalité 1 de T_OP_PREC :	TO_INCONNU TOIC	
				Modalité 2 de T_OP_PREC :	TO_SECT_VI TOGV	
				Modalité 3 de T_OP_PREC :	TO_GROUP TOGR	
				Modalité 4 de T_OP_PREC :	TO_GIE TOG	
				Modalité 5 de T_OP_PREC :	TO_GIE_EX TOGE	
				Modalité 6 de T_OP_PREC :	TO_GIE_PA TOGP	
				Modalité 7 de T_OP_PREC :	TO_GIE_DV TOGD	
				Modalité 8 de T_OP_PREC :	TO_EXP_PHV TOPR	
				Modalité 9 de T_OP_PREC :	TO_FOTER TOFO	
				Modalité 10 de T_OP_PREC :	TO_FEDE_OP TOFD	
				Modalité 11 de T_OP_PREC :	TO_AUTRE TOAU	

Structuration des informations

Début enquête OP



Fin enquête OP



## le 8. 7.94 à 19:16

\_\_\_\_\_

\_\_\_\_\_

Inconnu	[ 11
Section villageoise	[ 12
Exploitant privé	[ 13
Autre	[ 14
Groupe ment	[ 15
GIE	[ 16
Foyer	[ 17
Coopératives	[ 18
Autre :.....	

[illegible]

Oui [ ]1  
Non [ ]2

Si la réponse à la question 6 est égale à 2 (Non), allez à la question 9

*Inventaire du matériel*

<b>Question no 7 : Code du matériel</b>	<b>Question no 8 : Type du matériel (codez votre réponse) Tracteur = 1, Offset = 2, Moissonneuse batteuse = 3, Décortiqueuse = 4, GMP = 5</b>	<b>Question no 9 : Quantité</b>
_ _	[ ]	_ _
_ _	[ ]	_ _
_ _	[ ]	_ _
_ _	[ ]	_ _
_ _	[ ]	_ _
_ _	[ ]	_ _
_ _	[ ]	_ _
_ _	[ ]	_ _
_ _	[ ]	_ _
_ _	[ ]	_ _

*Enregistrer le matériel en état de marche ou facilement réparable*

**Question no 9 : Notes diverses sur l'OP**

.....

.....

.....

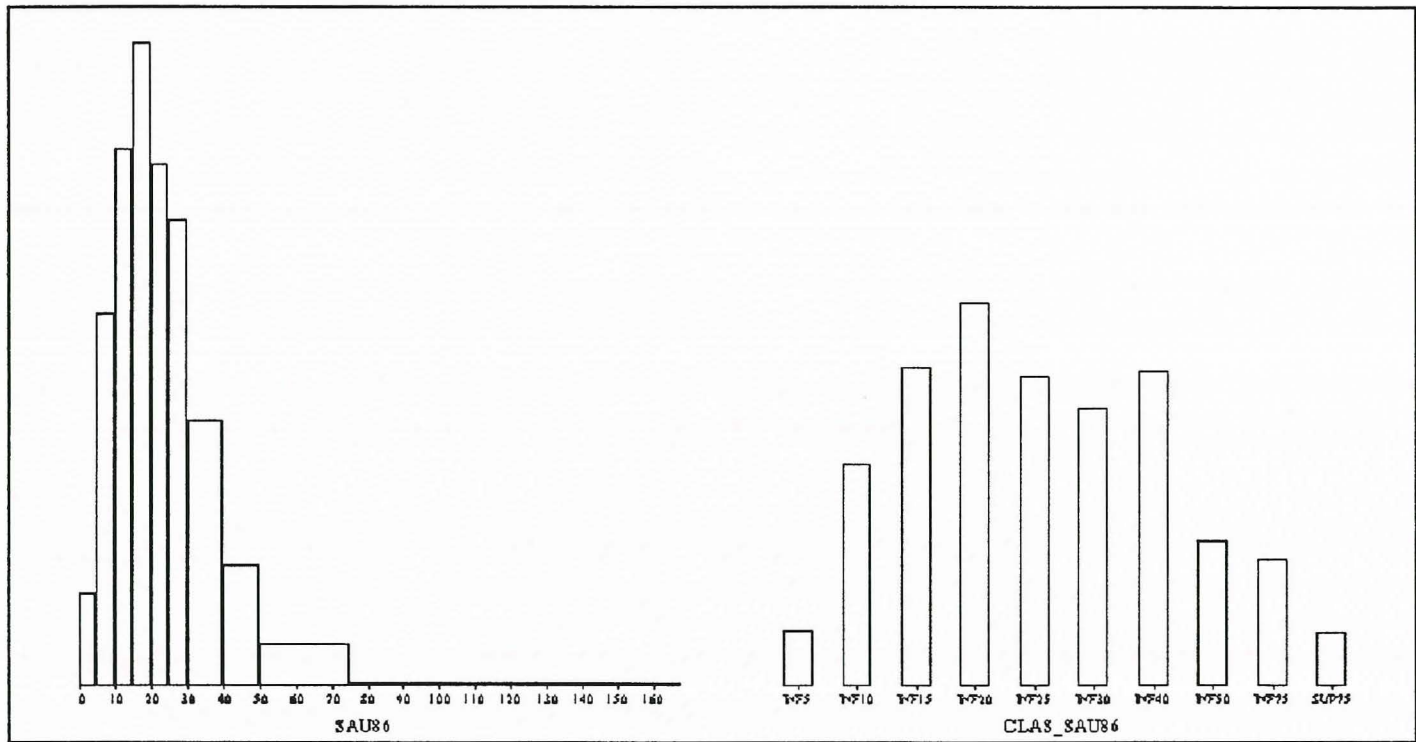
.....

.....

Indiquez ici vos commentaires

*Fin du questionnaire*

Annexe - 9 : Histogramme et diagramme en bâtons

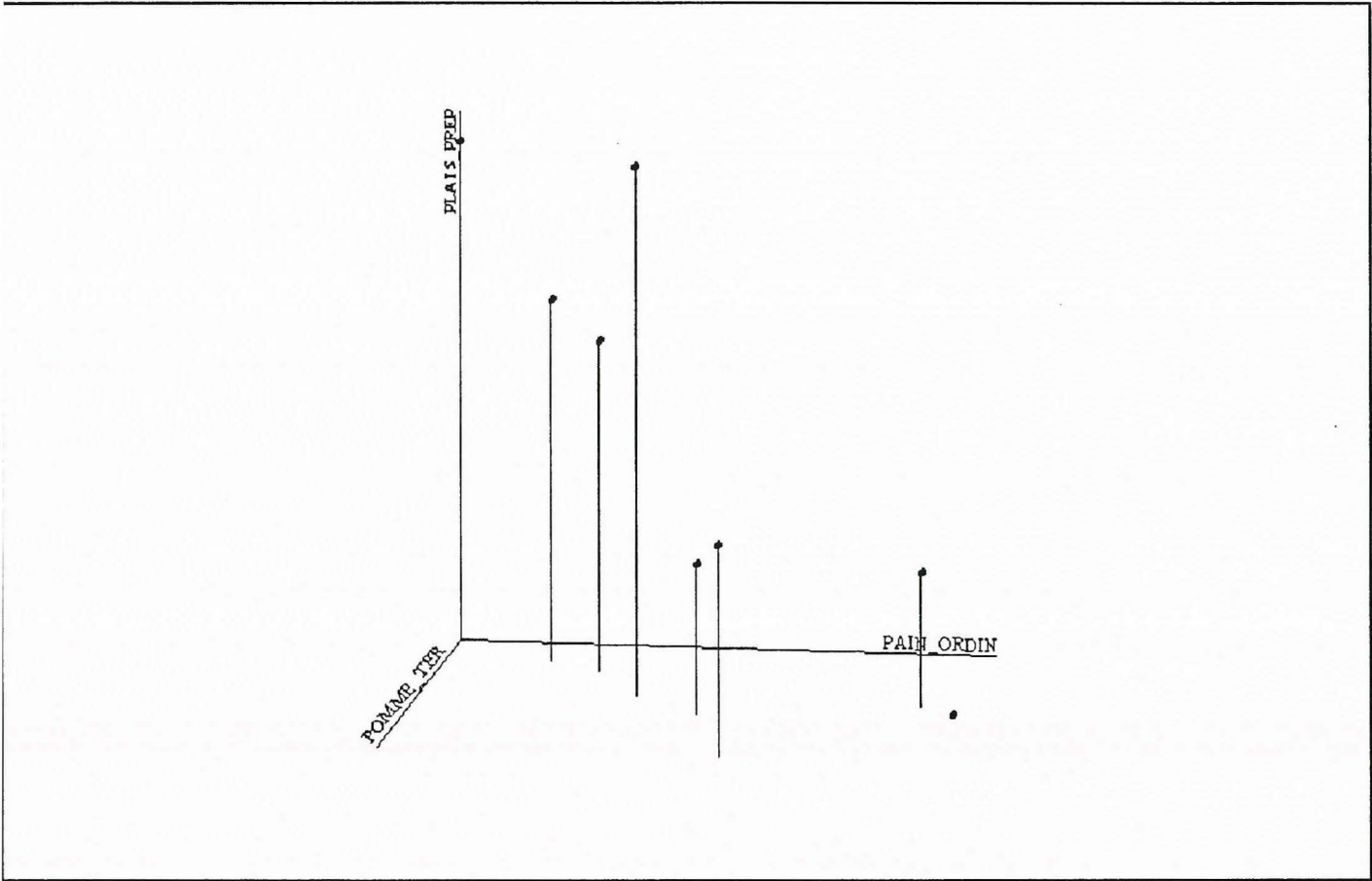


La même information est représentée sur ces deux graphiques : soit sous forme d’histogramme de la variable continue SAU86 (à gauche), soit de diagramme en bâtons de la variable classe de SAU86 (à droite).

La représentation de droite donne une fausse image de la distribution statistique des surfaces utiles d’exploitations en donnant une forme relativement symétrique qui est loin d’être la réalité.



Annexe - 10 : Nuage de points multidimensionnel



L'analyse en composantes principales permet de faire apparaître la structure, les lignes de force de nuages de points de dimension quelconque.

## **RESUME**

La SAED, pour répondre à sa mission d'observatoire du développement rural, doit renforcer sa capacité d'études, en particulier dans la maîtrise de l'outil statistique.

La présente mission réalisée avec l'appui du Ministère Français de la Coopération et du Développement, a été l'occasion de former des cadres de la SAED à la mise en oeuvre des méthodes statistiques dans le domaine des enquêtes en milieu rural.

-----

## **MOTS CLEFS**

SAED - FORMATION - STATISTIQUE UNI ET MULTI-DIMENSIONNELLE -  
LISA - TRAITEMENT D'ENQUETES